MapReduce Service

FAQs

Issue 01
Date 2020-07-16
Contents

1 Product Consultation.................................................................................................................. 1
  1.1 What Is MRS Used For?........................................................................................................... 1
  1.2 What Types of Distributed Storage Does MRS Support?............................................... 1
  1.3 How Do I Create an MRS Cluster Using a Custom Security Group?.......................... 1
  1.4 How do I Use MRS?.............................................................................................................. 1
  1.5 How Can I Ensure Data and Service Running Security?.................................................. 2
  1.6 Is a Real-time Data Warehouse Capable of Real-time Big Data Processing and Storage? 3
  1.7 What Are Regions and AZs?............................................................................................... 3
  1.8 How Do I Select a Region?.................................................................................................. 4
  1.9 How Do I Select an AZ?...................................................................................................... 4
  1.10 How Do I Obtain an OBS Endpoint?.................................................................................. 5
  1.11 How Can I Configure the Phoenix Connection Pool?.................................................... 5
  1.12 Does MRS Support Network Segment Change?............................................................ 5
  1.13 Can I Reduce the Configuration of an MRS Cluster Node?......................................... 5
  1.14 What Is the Relationship Between Hive and Other Components?.............................. 5
  1.15 Does an MRS Cluster Support Hive on Spark?.............................................................. 6
  1.16 Is the Current Hive Version Compatible with Another Version?................................. 6
  1.17 Which MRS Cluster Version Supports Hive Connection and User Synchronization?.... 6
  1.18 What Is the Relationship Between Spark and Hadoop?............................................... 6
  1.19 Does an MRS Cluster Support Spark on Hive?.............................................................. 6
  1.20 What Are the Differences Between OBS and HDFS in Data Storage?....................... 7
  1.21 How Do I Obtain the Hadoop Pressure Test Tool?......................................................... 7
  1.22 What Is the Relationship Between Impala and Other Components?.......................... 7
  1.23 What Is the Relationship Between Kudu and HBase?.................................................... 8
  1.24 Does an MRS Cluster Support Hive on Kudu?.............................................................. 8
  1.25 Solution for Scenarios with 1 Billion Data.................................................................... 8
  1.26 How Do I Change the IP address of DBService?........................................................... 8
  1.27 Can I Clear MRS sudo Logs?............................................................................................ 8

2 Billing................................................................................................................................. 10
  2.1 How Is the MRS Billed?...................................................................................................... 10
  2.2 How Does a Product Quotation Work When an MRS Cluster Is Created?..................... 10
  2.3 How Is the Auto Scaling Function Charged in an MRS Cluster?................................. 10
  2.4 Can the Billing Mode Be Changed From Pay-per-use to Yearly/Monthly?...................... 11
2.5 How Is Manual Cluster Scale-out Billed?.............................................................................................................. 11
2.6 How Is the MRS Renewed?........................................................................................................................................... 11
2.7 How Is the Task Node in an MRS Cluster Billed?....................................................................................................... 11
2.8 Why Is An Error Reported When I Unsubscribe From an ECS after Unsubscribing From MRS?.......................... 11
3 Account and Password.................................................................................................................................................... 12
3.1 What Is the Account for Logging In to MRS Manager?................................................................................................. 12
4 Accounts and Permissions................................................................................................................................................ 13
4.1 Does an MRS Cluster Support Access Permission Control If Kerberos Authentication Is not Enabled?................. 13
4.2 How Do I Assign Tenant Management Permission to a New Account?.................................................................... 13
4.3 Can an IAM User of MRS Be Associated with an MRS Manager User?................................................................. 14
4.4 How Do I Create an IAM Agency for an MRS Cluster?............................................................................................. 14
4.5 How Do I Set the Permission Granularity of MRS Manager and IAM to the Table or Database Level?.............. 14
4.6 How Do I Customize an MRS Policy?.......................................................................................................................... 14
4.7 Why Cannot I Find the Manage User Function on the System Page on MRS Manager?........................................ 15
4.8 What Are the Functions of IAM User Synchronization?............................................................................................ 15
4.9 Does Hue Support Account Permission Configuration?.............................................................................................. 15
5 Client Usage.................................................................................................................................................................. 16
5.1 How Do I Use a Component Client?............................................................................................................................. 16
5.2 Can I Download the Client If No EIP Is Bound?........................................................................................................... 16
5.3 Does a Flume Client Support Windows?...................................................................................................................... 16
6 Web Page Access............................................................................................................................................................. 17
6.1 How Do I Access MRS Manager From the Intranet?................................................................................................... 17
6.2 What Should I Do If an Alarm Indicating that the OBS Certificate Is About to Expire Is Displayed on MRS Manager? 17
6.3 How Do I Use HDFS When the Local IP Address and HUAWEI CLOUD IP Address Are in Different Network Segments?......................................................................................................................... 17
6.4 Changing the Session Timeout Interval of the Open Source Component Web UI...................................................... 17
6.5 Dynamic Resource Plan Page on the MRS Tenant Management Page Cannot Be Refreshed............................. 19
7 Alarm Monitoring............................................................................................................................................................. 20
7.1 In an MRS Streaming Cluster, Can the Kafka Topic Monitoring Send Alarms?....................................................... 20
8 Performance Tuning......................................................................................................................................................... 21
8.1 Does an MRS Cluster Support System Reinstallation?................................................................................................ 21
8.2 How Do I Improve the Resource Usage of Core Nodes in a Cluster?..................................................................... 21
8.3 How Do I Configure the Default Disk Usage of MRS?.................................................................................................. 21
8.4 How Do I Disable the Firewall Service to Prevent Impact on the System?............................................................... 21
8.5 How Do I Increase the HDFS Capacity of Nodes in an MRS Cluster Without Adding a Core Node?..................... 22
9 Job Development........................................................................................................................................................... 23
9.1 How Do I Prepare MRS Data Sources?.......................................................................................................................... 23
9.2 How Do I Submit a Program to MRS?............................................................................................................................ 24
9.3 How Do I Adjust the Heap Memory of the JobHistory Process? .......................................................... 24
9.4 What Types of Spark Jobs Can Be Submitted in a Cluster? .................................................................. 24
9.5 Can Only One Spark Task Run at a Time After the Minimum Tenant Resources of an MRS Cluster Is Changed to 0? ................................................................................................................................. 25

10 Cluster Upgrade/Patching ...................................................................................................................... 26
10.1 How Do I Upgrade MRS? ..................................................................................................................... 26
10.2 Can I Change the MRS Version? .......................................................................................................... 26
10.3 Can I Install Components in an MRS Cluster? ................................................................................... 26

11 Peripheral Ecosystem Interconnection .................................................................................................. 27
11.1 Can MRS Cluster Monitoring Metrics Be Integrated into Cloud Eye? .................................................. 27
11.2 Can MRS Be Used to Perform Read and Write Operations on DLI Service Tables? ........................................... 27
11.3 Can Tableau Be Interconnected with Hive in an MRS Cluster? .............................................................. 27
11.4 How Can I Access OBS from the Same AZ in the Same Region over the Intranet After MRS Is Established? Is There Any Speed Limitation? ........................................................................................................... 27
11.5 Can the Crawler Service Be Deployed in MRS? .................................................................................. 28
11.6 Does DWS and MRS Support Safe Deletion (Preventing Illegal Restoration After Deletion)? ................ 28
11.7 Does OBS Support the ListObjectsV2 Protocol? .................................................................................. 28
11.8 Can MRS Data Be Stored in a Parallel File System Provided by OBS? ................................................. 28
11.9 Can DLF Be Interconnected with MRS Clusters? .................................................................................. 28
11.10 Why Is the MRS Cluster with Kerberos Authentication Enabled Not Found When a Connection Is Set Up in DLF? ........................................................................................................................................ 28
11.11 How Do I Use PySpark on an ECS to Connect to an MRS Spark Cluster with Kerberos Authentication Enabled on the Intranet? ........................................................................................................ 29
11.12 Why Is There No Mapping Field When HBase Synchronizes Data to CSS? ........................................... 29
11.13 Can a Spark Cluster Access Data in OBS? ............................................................................................ 29
11.14 What Are the Differences Between Spark in MRS and Spark in DLI? ................................................ 29
11.15 How Can I Import Data from OBS to HDFS in Flume? ....................................................................... 29
11.16 Does the Component Interconnection Support the External KDC? ....................................................... 29
11.17 When the Open Source Kylin 3.x Is Interconnected with the MRS 1.9.3 Version, the Jetty Version Compatibility Problem Exists ........................................................................................................... 30

12 Cluster Access ....................................................................................................................................... 31
12.1 Can I Switch Between the two Login Modes of MRS? .......................................................................... 31
12.2 How Do I Remotely Use an MRS Cluster? .............................................................................................. 31
12.3 How Do I Set the IP Address and Port Number of the MRS Cluster for Local Joint Commissioning? ........ 31
12.4 How Do I Obtain the ZooKeeper Address? ............................................................................................ 31

13 Big Data Service Development .......................................................................................................... 34
13.1 What Are the Differences Between Sample Project Building and Application Development in MapReduce? Is Python Code Supported? .................................................................................................. 34
13.2 What Compression Algorithms Does HBase Support? ........................................................................... 34
13.3 How Do I Use HBase? ............................................................................................................................ 34
13.4 Can MRS Write Data to HBase Through the HBase External Table of Hive? ........................................ 34
13.5 How Do I View HBase Logs? .................................................................................................................. 35
13.6 How Do I Access HBase Shell in a Cluster with Kerberos Authentication Enabled?................................. 35
13.7 Which Number Type of DynamoDB Is Better in Hive Tables?........................................................................ 35
13.8 Can the Hive Driver Be Interconnected with DBCP2?.................................................................................. 35
13.9 How Does User A View the Hive Table Created by User B?........................................................................ 35
13.10 How Do I Use IDEA on a Local Computer to Access Hive?....................................................................... 35
13.11 How Do I Commission a SparkSQL Job in the Windows Environment?..................................................... 36
13.12 How Do I Reset Kafka Data?....................................................................................................................... 36
13.13 Can MRS Run Multiple Flume Tasks at the Same Time?............................................................................ 36
13.14 How Do I Change FlumeClient Logs to Standard Output Logs?................................................................. 36
13.15 How Does HDFS Balance Data?.................................................................................................................. 37
13.16 How Do I Change the Number of HDFS Replicas?...................................................................................... 37
13.17 How Do I Use Python to Remotely Connect to an HDFS Port?................................................................. 37
13.18 Where Are the JAR Files and Environment Variables of Hadoop Located?............................................. 38
13.19 What Compression Algorithms Does Kudu Support?.................................................................................. 38
13.20 How Do I View Kudu Logs?......................................................................................................................... 38
13.21 Does OpenTSDB Support Python APIs?...................................................................................................... 38
13.22 How Do I View Container Logs of Yarn?.................................................................................................... 38
13.23 How Do I Specify a Log Path When Submitting a Task in an MRS Storm Cluster?.................................. 39
13.24 Where Can I Download the Dependency Package (com.huawei.gaussc10) in the Hive Sample Project?........................................................................................................................................ 39
13.25 Does HDFS Support OBS?........................................................................................................................... 39
13.26 Why Does the Kafka Access Protocol Configuration Not Take Effect?....................................................... 39
13.27 How Do I Set and Modify the Data Retention Period of an HBase Table?.................................................... 39
13.28 How Do I Specify an Application Name in the flink run Command?.......................................................... 40
13.29 How Do I Connect to HBase of MRS Through HappyBase?..................................................................... 40
13.30 How Do I Change the Resource Ratio of Yarn Queues?............................................................................ 40
13.31 How Do I Specify a Flink Task to Run in a Specified Yarn Queue?.............................................................. 40
13.32 How DO I View the Client Version of MRS Kafka?....................................................................................... 40
13.33 What Access Protocols Are Supported by Kafka?...................................................................................... 40
13.34 How Do I Configure Other Data Sources on Presto?.................................................................................. 40
13.35 Does Hive Query Data Can Be Exported?.................................................................................................... 41
13.36 How Does MRS Connect to Impla Shell?..................................................................................................... 41
13.37 How Does MRS Connect to Spark Shell?...................................................................................................... 41
13.38 How Does MRS Connect to Spark Beeline?................................................................................................. 41
13.39 Where Are the Execution Logs of Spark Jobs Stored?............................................................................... 42
14 API...........................................................................................................................................43
14.1 How Do I Configure the node_id Parameter When Using the API for Adjusting Cluster Nodes?............. 43
15 Cluster Management........................................................................................................................................ 44
15.1 How Do I View All Clusters?........................................................................................................................ 44
15.2 How Do I View Log Information?.................................................................................................................. 44
15.3 How Do I View Configuration Information of a Cluster?.......................................................................... 45
15.4 How Do I Install Streaming Cluster Components in an MRS Analysis Cluster?.................................. 45
15.5 How do I Install Kafka and Flume in an MRS Cluster? ................................................................. 45
15.6 How Do I Stop an MRS Cluster? ........................................................................................................ 46
15.7 Do I Need to Shut Down a Master Node Before Scaling Up Its Specifications? .................. 46
15.8 Can the Capacity of a Data Disk Be Expanded on MRS? .............................................................. 46
15.9 How Do I Add Components to an Existing Cluster? .................................................................... 46
15.10 Can the Components Installed in an MRS Cluster Be Deleted? ................................................... 46
15.11 Can the Node Type of an MRS Cluster Be Changed? ................................................................. 46
15.12 Can MRS Cluster Nodes Be Changed on the MRS Management Console? .................. 46
15.13 Can Only Core and Task Nodes with the Same Configurations as Existing Cluster Nodes Be Added During Cluster Scale-out? ......................................................................................... 47
15.14 Can the MRS Cluster in Use Be Expanded? ........................................................................... 47
15.15 How Do I Cancel Cluster Risk Notification? .............................................................................. 47
15.16 How Do I Configure Parameters for Bootstrap Actions? ............................................................ 47

16 Kerberos Usage.......................................................................................................................... 48
16.1 Can I Change the Kerberos Authentication Status of a Purchased MRS Cluster? .................. 48
16.2 How Do I Access MRS Manager That Supports Kerberos Authentication? .......................... 48
16.3 What Are the Ports of the Kerberos Authentication Service? ...................................................... 48
16.4 How Do I Deploy the Kerberos Service in a Running Cluster? .................................................... 48
16.5 How Do I Access Hive in a Cluster with Kerberos Authentication Enabled? ...................... 49
16.6 How Do I Access Presto in a Cluster with Kerberos Authentication Enabled? .................... 49

17 Metadata Management.............................................................................................................. 51
17.1 Where Can I View Hive Metadata? ......................................................................................... 51
1 Product Consultation

1.1 What Is MRS Used For?

Based on the Hadoop open source software, Spark in-memory computing engine, HBase distributed storage database, and Hive data warehouse framework, MRS provides a unified platform for enterprise-level big data storage, query, and analysis. Using MRS, you can quickly establish a system to process massive amounts of data and solve the following problems:

- Analysis and computing of massive amounts of data
- Storage of massive amounts of data
- Stream processing of massive amounts of data

1.2 What Types of Distributed Storage Does MRS Support?

MRS currently supports Hadoop 3.1.x and will support other mainstream Hadoop versions released by the community. For details about the component versions supported by MRS, see List of MRS Component Versions.

1.3 How Do I Create an MRS Cluster Using a Custom Security Group?

If you want to use a security group created by yourself when buying a cluster, you need to enable ports such as 9022. Alternatively, select Auto create in Security Group when buying a cluster.

1.4 How do I Use MRS?

MapReduce Service (MRS) is a HUAWEI CLOUD service that is used to deploy and manage the Hadoop system and enables one-click Hadoop cluster deployment. MRS provides enterprise-level big data clusters on the cloud. Tenants can fully
control the clusters and easily run big data components such as Hadoop, Spark, HBase, Kafka, and Storm in the clusters.

MRS is easy to use. You can execute various tasks and process or store PB-level data using computers connected in a cluster. The procedure of using MRS is as follows:

1. Develop a data processing program. **MRS Developer Guide** provides sample code and tutorials to help you quickly develop your own programs and properly run them.
2. Upload local programs and data files to OBS.
3. Create a cluster, choose a cluster type for offline data analysis or stream processing or both, and set ECS instance specifications, instance count, data disk type (common I/O, high I/O, and ultra-high I/O), and components to be installed such as Hadoop, Spark, HBase, Hive, Kafka, and Storm in a cluster. You can use a bootstrap action to execute a script on a specified node before or after the cluster is started to install additional third-party software, modify the cluster running environment, and perform other customizations.
4. Manage a job. MRS provides a platform for executing programs you develop. You can submit, execute, and monitor such programs on MRS.
5. Manage a cluster. MRS provides you with MRS Manager, an enterprise-level unified management platform of big data clusters, helping you quickly know health status of services and hosts. Through graphical metric monitoring and customization, you can obtain critical system information in a timely manner. In addition, you can modify service attribute configurations based on service performance requirements, and start or stop clusters, services, and role instances in one click.
6. Terminate a cluster. You can terminate an MRS cluster that is no longer used after job execution is complete. The terminated cluster is no longer billed.

### 1.5 How Can I Ensure Data and Service Running Security?

MRS is a platform for massive data management and analysis and has high security. It provides multiple features to ensure secure running of data and services:

- Network isolation
  The entire public cloud network is divided into two planes: service plane and management plane. The two planes are deployed and isolated physically to ensure the security of the service and management networks.
  - Service plane: network plane where cluster components are running. It provides service channels for storing and accessing data, submitting tasks, and computing data.
  - Management plane: public cloud management console on which you can purchase and manage MRS.

- Host security
  Users can deploy third-party antivirus software based on their service requirements. MRS provides the following measures to improve the security of OSs and ports:
- Security hardening of OS kernels
- OS patch update
- OS permission control
- OS port management
- OS protocol and port attack defense

- Data security
  MRS allows data to be stored on OBS, ensuring data security.
- Data integrity
  After processing data, MRS transmits the data to OBS through SSL, ensuring data integrity.

1.6 Is a Real-time Data Warehouse Capable of Real-time Big Data Processing and Storage?

MRS can meet this requirement. For details, see Application Scenarios.

1.7 What Are Regions and AZs?

A region and availability zone (AZ) identify the location of a data center. You can create resources in a specific region and AZ.

- Regions are defined in terms of geographical locations and network latencies. Public services such as Elastic Cloud Server (ECS), Elastic Volume Service (EVS), Object Storage Service (OBS), Virtual Private Cloud (VPC), Elastic IP (EIP), and Image Management Service (IMS) are shared within the same region. Regions are classified as common regions and dedicated regions. A common region provides common cloud services that should be made available to all tenants. A dedicated region provides services of a specific type or only for specific tenants.

- An availability zone (AZ) contains one or multiple physical data centers. Each AZ has independent cooling, fire extinguishing, moisture-proof, and electricity facilities. Within an AZ, computing, network, storage, and other resources are logically divided into multiple clusters. AZs within a region are interconnected using high-speed optical fibers to allow you to build cross-AZ high-availability systems.

Figure 1-1 shows the relationship between regions and AZs.
1.8 How Do I Select a Region?

When selecting a region, consider the following factors:

- **Location**
  You are advised to select a region close to you or your target users. This reduces network latency and improves access speed. However, the Chinese mainland regions provide the same infrastructure, BGP network quality, as well as resource operations and configurations. Therefore, if you or your target users are in the Chinese mainland, you do not need to consider the network latency differences when selecting a region.
  - If you or your target users are in the Asia Pacific area (excluding the Chinese mainland), select the **AP-Hong Kong**, **AP-Bangkok**, or **AP-Singapore** region.
  - If you or your target users are in Africa, select the **AF-Johannesburg** region.
  - If you or your target users are in Europe, select the **EU-Paris** region.

- **Resource Price**
  Resource prices may vary in different regions. For details, see [Product Pricing Details](#).

1.9 How Do I Select an AZ?

When determining whether to deploy resources in the same AZ, consider your applications’ requirements for disaster recovery (DR) and network latency.

- For high DR capability, deploy resources in different AZs in the same region.
- For low network latency, deploy resources in the same AZ.
1.10 How Do I Obtain an OBS Endpoint?

When using an API to use resources, you must specify its region and endpoint. For details about HUAWEI CLOUD regions and endpoints, see Regions and Endpoints.

1.11 How Can I Configure the Phoenix Connection Pool?

Phoenix does not support connection pool configuration. You are advised to write code to implement a tool class for managing connections and simulate a connection pool. For details, see https://stackoverflow.com/questions/35183713/what-is-the-correct-way-to-pool-the-phoenix-query-server-connections.

1.12 Does MRS Support Network Segment Change?

MRS does not support network segment change.

1.13 Can I Reduce the Configuration of an MRS Cluster Node?

Currently, you cannot reduce the configuration of an MRS cluster node. If you have such a requirement, you are advised to submit a request.

1.14 What Is the Relationship Between Hive and Other Components?

- Relationship Between Hive and HDFS
  Hive is a sub-project of Apache Hadoop, which uses HDFS as the file storage system. It parses and processes structured data with highly reliable underlying storage supported by HDFS. All data files in the Hive database are stored in HDFS, and all data operations on Hive are also performed using HDFS APIs.

- Relationship Between Hive and MapReduce
  Hive data computing depends on MapReduce. MapReduce is also a sub-project of Apache Hadoop and is a parallel computing framework based on HDFS. During data analysis, Hive parses HiveQL statements submitted by users into MapReduce tasks and submits the tasks for MapReduce to execute.

- Relationship Between Hive and DBService
  MetaStore (metadata service) of Hive processes the structure and attribute information about Hive databases, tables, and partitions. The information needs to be stored in a relational database and is maintained and processed by MetaStore. In MRS, the relational database is maintained by the DBService component.
Relationship Between Hive and Spark
Hive data computing can also be implemented on Spark. Spark is an Apache project and is an in-memory distributed computing framework. During data analysis, Hive translates HiveQL statements submitted by users into Spark tasks and submits the tasks for Spark to execute.

1.15 Does an MRS Cluster Support Hive on Spark?
Currently, only MRS 1.9.x clusters support Hive on Spark.

1.16 Is the Current Hive Version Compatible with Another Version?
Compared with Hive 1.2, Hive 3.1 is incompatible with the following:
- Field type constraint: String cannot be converted to INT.
- UDF incompatibility: The date type in UDF is changed to the built-in date type of Hive.
- The index function is discarded.
- Time function difference: Hive 3.1 uses the UTC time, and Hive 1.2 uses the local time zone.
- Driver incompatibility: The JDBC drivers of Hive 3.1 and Hive 1.2 are incompatible.
- In Hive 3.1, column names in ORC files are case sensitive and underscores (_) are sensitive.
- The time column cannot exist in the table of Hive 3.1.

1.17 Which MRS Cluster Version Supports Hive Connection and User Synchronization?
MRS 2.0.5 or later supports Hive connections on DAYU and provides the IAM user synchronization function.

1.18 What Is the Relationship Between Spark and Hadoop?
Spark is a fast and universal processing engine that is compatible with Hadoop data. Spark can run in a Hadoop cluster using Yarn. Additionally, Spark can process different types of data stored in HDFS, HBase, Hive, and Hadoop.

1.19 Does an MRS Cluster Support Spark on Hive?
MRS supports Spark on Hive. For details about Spark on Hive, see Reading Data from Hive and Writing Data to HBase.
1.20 What Are the Differences Between OBS and HDFS in Data Storage?

Data sources processed by MRS are from OBS or HDFS. OBS is an object-based storage service that provides you with massive, secure, reliable, and cost-effective data storage capabilities. MRS can process data in OBS directly. You can view, manage, and use data by using the web page of the cloud management platform or OBS Client. In addition, you can use REST APIs independently or integrate APIs to service applications to manage and access data.

- Data stored on OBS: Data storage and computing are performed separately. Cluster storage costs are low, and storage capacity is not limited. Clusters can be deleted at any time. The computing performance is determined by the OBS access performance and is lower than that of HDFS. OBS is recommended when data computing is infrequent.

- Data stored in HDFS: Data storage and computing are performed together. Cluster storage costs are high, and storage capacity is limited. The computing performance is high. Data must be exported and stored before the clusters are deleted. HDFS is recommended when data computing is frequent.

1.21 How Do I Obtain the Hadoop Pressure Test Tool?

Download it from the following URL: [https://github.com/Intel-bigdata/HiBench](https://github.com/Intel-bigdata/HiBench).

1.22 What Is the Relationship Between Impala and Other Components?

- Relationship Between Impala and HDFS
  Impala uses the HDFS as its file storage system. It parses and processes structured data with highly reliable underlying storage supported by Hadoop HDFS. Impala will not move data in HDFS and will provide faster access.

- Relationship Between Impala and Hive
  Impala uses Hive metadata, ODBC driver, and SQL syntax. Unlike Hive, Impala is not based on the MapReduce algorithm. Instead, it implements a distributed architecture based on daemon and is responsible for all aspects of query execution running on the same node. Therefore, it reduces the latency of using MapReduce, which makes Impala faster than Hive.

- Relationship Between Impala and MapReduce
  None

- Relationship Between Impala and Spark
  None

- Relationship Between Impala and Kudu
  Kudu is closely integrated with Impala to replace the combination of Impala, HDFS, and Parquet, allowing you to insert, query, update, and delete data from Kudu tablets using Impala's SQL syntax. In addition, you can use JDBC or ODBC. Impala functions as a proxy to connect to Kudu for data operations.
• Relationship Between Impala and HBase
  The default Impala table uses data files stored in HDFS, which is ideal for batch loading and query of full table scanning. However, HBase can provide convenient and efficient query of OLTP-style organization data.

1.23 What Is the Relationship Between Kudu and HBase?

Kudu is designed based on the HBase structure and can implement fast random read/write and update functions that HBase features. The two components are similar in architecture, but still have the following main differences:

- Kudu does not depend on ZooKeeper. It implements Raft to ensure consistency.
- The Kudu persistence data does not depend on HDFS. TServer implements strong data consistency and reliability.

1.24 Does an MRS Cluster Support Hive on Kudu?

MRS does not support Hive on Kudu.

Currently, MRS supports only the following two methods to access Kudu:

- Access Kudu through the Impala table.
- Use the client application to access and operate the Kudu table.

1.25 Solution for Scenarios with 1 Billion Data

- GaussDB (for MySQL) is recommended for scenarios such as data update, online transaction processing (OLTP), and complex analysis.
- The Impala and Kudu databases of MRS can also meet this requirement. The Impala and Kudu databases can load all join tables to the memory during the join operation.

1.26 How Do I Change the IP address of DBService?

MRS does not support the change of the DBService IP address.

1.27 Can I Clear MRS sudo Logs?

MRS sudo log files are operation records of user omm and can be cleared to facilitate fault locating. Logs occupy a part of the storage space. Therefore, you are advised to delete operation logs that are generated a long time ago to release the storage space.

1. If the log file size is large, you can add the log file directory to /etc/logrotate.d/syslog to enable the system to age logs and periodically delete logs that have been stored for a long time.
Method: Change the file log directory to sed -i '3 a/var/log/sudo/sudo.log' /etc/logrotate.d/syslog.

2. You can set /etc/logrotate.d/syslog based on the number and size of logs. If the number or size of logs exceeds the threshold, the logs are automatically deleted. By default, logs are aged based on the size and number of archived logs. You can use size and rotate to limit the size and number of archived logs, respectively. By default, there is no time limit. If you need to set the time, you can add daily/weely/monthly to specify the period.
2.1 How Is the MRS Billed?

MRS supports two billing modes: pay-per-use and yearly/monthly. For billing details, see Price Calculator.

- Yearly or monthly subscription: You can pay for clusters by year or month. The service use duration ranges from one month to one year.
- Pay-per-use: Nodes are charged by actual duration of use, with a billing cycle of one hour.

**NOTE**

- Clusters in the Starting, Failed, and Terminated state are not charged, but clusters in other states are charged.
- The fee here contains only the cost on clusters. The cost on data storage, bandwidth, and traffic on MRS are excluded.
- You will be notified of renewal if there is no sufficient balance for fee deduction. Cluster resources will be frozen during a retention period and unfrozen after your renewal.

2.2 How Does a Product Quotation Work When an MRS Cluster Is Created?

A product quotation displayed at the bottom of the page does not include disk fee if you specify only the number of disks without configuring instance count, and will include disk fee after you specify both the number of disks and instance count.

2.3 How Is the Auto Scaling Function Charged in an MRS Cluster?

A product quotation displayed at the bottom of the page does not include an auto scaling fee if you specify only the node range without configuring instance count, and will include the auto scaling fee after you configure both the node range and instance count.
If nodes are added through the auto scaling function when the cluster is in use, the added nodes are charged by actual usage duration with a one-hour billing cycle regardless of whether the cluster is charged in yearly/monthly or pay-per-use mode.

2.4 Can the Billing Mode Be Changed From Pay-per-use to Yearly/Monthly?

MRS does not support billing mode change. You can reduce costs by shutting down ECSs.

2.5 How Is Manual Cluster Scale-out Billed?

The billing mode of the manual cluster scale-out is the same as that of existing nodes of the same type. The billing is based on the resources used by the added nodes plus existing nodes. A new order is generated during the manual cluster scale-out. Billing starts only after the scale-out is successful. For details, see Manually Scaling Out a Cluster.

2.6 How Is the MRS Renewed?

MRS provides two billing modes, pay per use and yearly or monthly subscription. In pay per use mode, fees are deducted every hour and an insufficient balance may cause overdue payments. For yearly or monthly subscription, clusters need to be renewed before they expire. If your subscription is not renewed, your services will keep running, but enter into a retention period, during which MRS clusters will stop running but data is retained.

To renew the subscription, go to the Renewals page.

2.7 How Is the Task Node in an MRS Cluster Billed?

Task nodes in a yearly/monthly or pay-per-use cluster are billed in pay-per-use mode. That is, they are billed based on the actual usage duration by hours.

2.8 Why Is An Error Reported When I Unsubscribe From an ECS after Unsubscribing From MRS?

1. In the MRS cluster in use, query the ID of the ECS and ensure that the ECS is not in use.
2. On the ECS console, locate the ECS to be unsubscribed from and click Locked by MRS to unlock the ECS.
3. Click Unsubscribe again.
4. If the unsubscription exception persists, collect the ECS ID and contact HUAWEI CLOUD technical support.
3 Account and Password

3.1 What Is the Account for Logging In to MRS Manager?

The default account for logging in to MRS Manager is `admin`, and the password is set during cluster creation.
4.1 Does an MRS Cluster Support Access Permission Control If Kerberos Authentication Is not Enabled?

For versions earlier than MRS 1.8.0, MRS clusters without Kerberos authentication enabled do not support access permission control. For versions earlier than MRS 1.8.0, the role management permission is available only when Kerberos authentication is enabled. In MRS 1.8.0 and later versions, all clusters have the role management permission.

On MRS Manager, choose System > Configuration > Permission to query the permission.

For details on how to configure permissions, see Permission Management.

For details on how to use clusters with Kerberos enabled, see Clusters with Kerberos Authentication Enabled.

4.2 How Do I Assign Tenant Management Permission to a New Account?

You can assign tenant management permission only in analysis and hybrid clusters and not in streaming clusters. Perform the following operations to assign tenant management permission to a new account:

Log in to MRS Manager as user admin. Choose System > Manage User and select the new account. Click Modify in the Operation column. Click Select and Add Role in Assign Rights by Role. If the Manager_tenant role is bound to an account, the account has permission to view the tenant management. If the Manager_administrator role is bound to an account, the account has permission to view and operate the tenant management.
4.3 Can an IAM User of MRS Be Associated with an MRS Manager User?

1. IAM users cannot be associated with MRS Manager users, but can be synchronized to MRS Manager. For details, see Synchronizing IAM Users to MRS.

2. If you want IAM user A to have the permissions of MRS Manager user B, log in to MRS Manager, choose System > Manage User, and bind user A to the same role as user B.

4.4 How Do I Create an IAM Agency for an MRS Cluster?

HUAWEI CLOUD services are interacted with each other. If some cloud services need collaborative work, you need to create a cloud service agency to delegate other cloud services to perform resource O&M on behalf of you. For details, see Cloud Service Delegation. For details about how to configure the fine-grained permission for multiple MRS users to access OBS, see Configuring Fine-Grained Permissions for MRS Multi-User Access to OBS.

4.5 How Do I Set the Permission Granularity of MRS Manager and IAM to the Table or Database Level?

IAM users cannot be associated with MRS Manager users, but can be synchronized to MRS Manager. For details, see Synchronizing IAM Users to MRS.

Table- and database-level permissions of users can be added only on MRS Manager. For details, see Configuring Hive Permissions.

4.6 How Do I Customize an MRS Policy?

1. On the IAM console, choose Permissions in the navigation pane. Then click Create Custom Policy.

2. Specify Policy Name for the custom policy.

3. Set Scope to Project-level service.

4. Specify Policy View. The following options are supported:
   - Visual editor: Customize a policy as required by selecting cloud services, actions, resources, and request conditions. This does not require knowledge of JSON syntax.
   - JSON: Edit JSON policies from scratch or based on an existing policy.

You can also click Select Existing Policy/Role in the Policy Content area to select an existing policy as the template for modification.

5. (Optional) Enter a brief description in the Description area.
6. Click **OK**.
7. Assign the policy to a user group so that users in the group can inherit the permissions of the policy.

**NOTE**

For details, see *Creating a Custom Policy*.

### 4.7 Why Cannot I Find the Manage User Function on the System Page on MRS Manager?

You do not have the **Manager_administrator** role permission. Therefore, **Manage User** is not displayed on the **System** page of MRS Manager.

### 4.8 What Are the Functions of IAM User Synchronization?

IAM user synchronization is to synchronize IAM users bound with MRS policies to the MRS system and create accounts sharing same names with the IAM users but different passwords. Then, you can use an IAM username (the password needs to be reset by user **admin** of MRS Manager) to log in to MRS Manager for cluster management, and submit jobs on the GUI in a cluster with Kerberos authentication enabled. For details, see *Synchronizing IAM Users to MRS*.

### 4.9 Does Hue Support Account Permission Configuration?

Hue does not provide the function of configuring account permissions. You can configure account permissions by configuring user roles and user groups on the **System** tab on MRS Manager.
5 Client Usage

5.1 How Do I Use a Component Client?

1. Log in to any Master node as user root.
2. Run the su omm command to switch to user omm.
3. Run the cd /opt/client command to switch to the client.
4. Run the source bigdata_env command to configure environment variables.
   If Kerberos authentication is enabled for the current cluster, run the kinit admin command to authenticate the user. If Kerberos authentication is disabled, skip this step.
5. After the environment variables are configured, run the client command of the component. For example, to view component information, you can run the HDFS client command hdfs dfs -ls / to view the HDFS root directory file.

5.2 Can I Download the Client If No EIP Is Bound?

If you have not applied for an EIP, you can download and update the MRS client. For details, see Using an MRS Client on Nodes Outside a Cluster.

5.3 Does a Flume Client Support Windows?

A Flume client cannot be installed on a Windows host. If you want to use the Flume client on another node, ensure that the node and MRS cluster are in the same VPC and subnet. For details, see Using an MRS Client on Nodes Outside a Cluster.
6.1 How Do I Access MRS Manager From the Intranet?

If your computer and the MRS cluster are not on the same intranet, you can create a Windows ECS on the ECS console to access MRS Manager. For details, see Access Using a Windows ECS.

6.2 What Should I Do If an Alarm Indicating that the OBS Certificate Is About to Expire Is Displayed on MRS Manager?

For details, see How Do I Handle the Expired OBS Certificate in the Cluster?.

Certificates on different nodes vary. Therefore, you need to perform the operations on each node separately. You are not advised to update the certificate on one node and copy the certificate to all nodes.

6.3 How Do I Use HDFS When the Local IP Address and HUAWEI CLOUD IP Address Are in Different Network Segments?

You can bind an EIP to the cluster and use the EIP to access HDFS of the cluster. For details, see EIP-based Access.

6.4 Changing the Session Timeout Interval of the Open Source Component Web UI

You are advised to set a proper web page timeout interval to prevent information leakage caused by long-time web page exposure.
On the cluster details page, choose **Components > meta > Service Configuration**, switch **Basic** to **All**, and search for the **http.server.session.timeout.secs**. If the configuration item exists, perform the following steps to modify it. If this configuration item does not exist, the version does not support dynamic adjustment of the session duration.

All timeout intervals must be set to the same value to avoid the conflict between the actual timeout interval and the configured timeout interval.

- **Changing the Timeout Intervals of the MRS Manager Page and Authentication Center Page**
  a. Log in to all Master nodes in the cluster and run the following command on each Master node:
  b. Change the value of `<session-timeout>20</session-timeout>` in the `/opt/Bigdata/apache-tomcat-7.0.78/webapps/cas/WEB-INF/web.xml` file. In the parameter, 20 indicates the session timeout interval, in minutes. Change it based on the site requirements. The maximum timeout interval is 480 minutes.
  c. Change the value of `<session-timeout>20</session-timeout>` in the `/opt/Bigdata/apache-tomcat-7.0.78/webapps/web/WEB-INF/web.xml` file. In the file, 20 indicates the session timeout interval, in minutes. Change it based on the site requirements. The maximum timeout interval is 480 minutes.
  d. Change the values of `p:maxTimeToLiveInSeconds="$tgt.maxTimeToLiveInSeconds:1200"` and `p:timeToKillInSeconds="$tgt.timeToKillInSeconds:1200"` in the `/opt/Bigdata/apache-tomcat-7.0.78/webapps/cas/WEB-INF/spring-configuration/ticketExpirationPolicies.xml` file. The value 1200 indicates the validity period of the authentication center, in seconds. The validity period cannot exceed 28,800 seconds.
  e. Restart the Tomcat node on the active Master node.
     i. On the active Master node, run the `netstat -anp | grep 28443 | grep LISTEN | awk '{print $7}'` command as user `omm` to query the Tomcat process ID.
     ii. Run the `kill -9 {pid}` command, in which `{pid}` indicates the Tomcat process ID obtained in the previous step.
     iii. Wait until the process automatically restarts. You can run the `netstat -anp | grep 28443 | grep LISTEN` command to check whether the process is successfully restarted. If the process can be queried, the process is successfully restarted. If the process cannot be queried, query the process again later.

- **Changing the Timeout Interval of an Open Source Component Web UI**
  a. On the cluster details page, choose **Components > meta > Service Configuration**.
  b. Change the value of **http.server.session.timeout.secs** under **meta** as required. The unit is second.
  c. Click **Save Configuration**, deselect **Restart the affected services or instances**, and click **OK**.

You are advised to perform the restart during off-peak hours.
d. (Optional) To use the Spark web page, choose **Components > Spark > Service Configuration**, switch **Basic** to **All**, and search for and change the value of **spark.session.maxAge** to a proper value. The unit is second.

e. Click **Save Configuration**, deselect **Restart the affected services or instances**, and click **OK**.

f. Restart the meta service and components on web UI, or restart the cluster during off-peak hours.

The restart affects services. You are advised to restart the service during off-peak hours or use the rolling restart function to restart the service without affecting services. For details, see **Performing Rolling Restart**.

### 6.5 Dynamic Resource Plan Page on the MRS Tenant Management Page Cannot Be Refreshed

1. Log in to the Master1 and Master2 nodes as user **root**.
2. Run the `ps -ef |grep aos` command to check the AOS process ID.
3. Run the `kill -9 AOS process ID` command to end the AOS process.
4. Wait until the AOS process is automatically restarted. You can run the `ps -ef | grep aos` command to check whether the AOS process exists. If the process exists, the restart is successful. If the process does not exist, retry later.
7.1 In an MRS Streaming Cluster, Can the Kafka Topic Monitoring Send Alarms?

The Kafka topic monitoring cannot send alarms by email or SMS message. Currently, you can view alarm information on MRS Manager.
8 Performance Tuning

8.1 Does an MRS Cluster Support System Reinstallation?

An MRS cluster does not support system reinstallation.

8.2 How Do I Improve the Resource Usage of Core Nodes in a Cluster?

1. Log in to MRS Manager.
2. Choose Services > Yarn > Service Configuration, and set Type to All.
3. Search for and change the value of yarn.nodemanager.resource.memory-mb. Increase the value based on the actual memory of the cluster node.
4. Click Save Configuration.
5. Select Restart the affected services or instances.
6. Click Yes.

8.3 How Do I Configure the Default Disk Usage of MRS?

The default disk usage of MapReduce is 90%. To change the disk usage, log in to MRS Manager, choose System > Configure Alarm Threshold, and change the alarm threshold.

8.4 How Do I Disable the Firewall Service to Prevent Impact on the System?

Step 1 Log in to each node of a cluster as user root.
Step 2 Run the `service iptables status` command to check whether the firewall service is enabled. If the firewall is enabled, go to Step 3 to disable the firewall.

Step 3 Run the `service iptables stop` command to stop the firewall service.

----End

8.5 How Do I Increase the HDFS Capacity of Nodes in an MRS Cluster Without Adding a Core Node?

If no Core node is added, the HDFS capacity of the nodes cannot be increased.

The capacity of HDFS is allocated by the cluster and cannot be greater than the data disk space of Core nodes. The Core node specifications have been determined during cluster creation. The disk space of the nodes cannot be changed after the cluster has been created. You can manually scale out the cluster to increase the capacity of HDFS. For details, see Manually Scaling Out a Cluster.
9 Job Development

9.1 How Do I Prepare MRS Data Sources?

MRS can process data in both OBS and HDFS. You need to prepare data before using MRS to analyze it.

1. Upload local data to OBS.
   a. Log in to the OBS management console.
   b. Create the **userdata** bucket on OBS and create the **program**, **input**, **output**, and **log** folders in the **userdata** bucket.
      i. Click **Create Bucket** to create a bucket named **userdata**.
      ii. In the OBS bucket list, click the bucket name **userdata**, choose **Objects > Create Folder**, and create the **program**, **input**, **output**, and **log** directories.
   c. Upload data to the **userdata** bucket.
      i. Go to the **program** folder and click **Upload Object**.
      ii. Click **add file** and select a user program.
      iii. Click **Upload**.
      iv. Upload the user data file to the **input** directory using the same method.

2. Import OBS data to HDFS.
   You can import OBS data to HDFS only when **Kerberos Authentication** is disabled and the cluster is running.
   a. Log in to the MRS management console.
   b. Click the name of the cluster to enter its details page.
   c. Go to the **Files** tab page and click **HDFS File List**.
   d. Go to the data storage directory, for example, **bd_app1**.
      The **bd_app1** directory is only an example. You can use any directory on the page or create a new one.
   e. Click **Import Data** and click **Browse** to select an OBS path and an HDFS path. See **Figure 9-1**.
9.2 How Do I Submit a Program to MRS?

MRS provides a platform for executing programs you have developed. You can submit, execute, and monitor such programs on MRS. How to submit a program to MRS? When adding a job, set Program Path to the path for storing programs. For details, see Managing Jobs.

9.3 How Do I Adjust the Heap Memory of the JobHistory Process?

Log in to MRS Manager and choose Services > Spark > Service Configuration > All > JobHistory > Default. Increase the value of the SPARK_DAEMON_MEMORY parameter as required. For details, see ALM-43006 Heap Memory Usage of the JobHistory Process Exceeds the Threshold.

9.4 What Types of Spark Jobs Can Be Submitted in a Cluster?

An MRS cluster supports Spark jobs submitted in Spark, Spark Script, or Spark SQL mode.
9.5 Can Only One Spark Task Run at a Time After the Minimum Tenant Resources of an MRS Cluster Is Changed to 0?

Only one Spark task can run at a time after the minimum tenant resources of an MRS cluster is changed to 0.
10 Cluster Upgrade/Patching

10.1 How Do I Upgrade MRS?
Currently, MRS cannot be smoothly upgraded from an earlier version to a later version. Currently, you can only purchase a cluster of the new version and migrate data from the cluster of the earlier version to the cluster of the latest version.

10.2 Can I Change the MRS Version?
The MRS version cannot be changed. You are advised to purchase an MRS cluster of a new version after unsubscribing from the cluster.

10.3 Can I Install Components in an MRS Cluster?
MRS does not allow you to upgrade MRS components. If you need components of a later version, purchase an MRS cluster of the corresponding version. For details about MRS component versions, see List of MRS Component Versions.
11 Peripheral Ecosystem Interconnection

11.1 Can MRS Cluster Monitoring Metrics Be Integrated into Cloud Eye?
No. The monitoring metrics within a cluster will not be integrated into Cloud Eye, because they belong to internal monitoring and are private customer information.

11.2 Can MRS Be Used to Perform Read and Write Operations on DLI Service Tables?
If data is stored on OBS, Spark in MRS can be used to read the DLI table, flexibly process the table data, and save the result to another DLI table. If data is not stored on OBS, Spark cannot be used to read the DLI table.

11.3 Can Tableau Be Interconnected with Hive in an MRS Cluster?
Tableau cannot be interconnected with Hive in an MRS cluster.

11.4 How Can I Access OBS from the Same AZ in the Same Region over the Intranet After MRS Is Established? Is There Any Speed Limitation?
1. After an MRS cluster is created, users in the same region and AZ can access OBS over intranet.
2. MRS and OBS have no limitations on either upload or download speed. When you access OBS over the public network, the speed is limited by the public network bandwidth. When you access OBS over intranet, the access is
11.5 Can the Crawler Service Be Deployed in MRS?

The crawler service cannot be deployed in MRS. You are advised to purchase an ECS to deploy the crawler service.

11.6 Does DWS and MRS Support Safe Deletion (Preventing Illegal Restoration After Deletion)?

Currently, MRS supports only restoration of backup data. DWS data can only be restored using snapshots. That is, both DWS and MRS support secure deletion.

11.7 Does OBS Support the ListObjectsV2 Protocol?

OBS does not support the ListObjectsV2 protocol.

11.8 Can MRS Data Be Stored in a Parallel File System Provided by OBS?

MRS supports the storage a parallel file system provided by OBS.

11.9 Can DLF Be Interconnected with MRS Clusters?

Data Lake Factory (DLF) can interconnect with MRS clusters through the V2 APIs. If the MRS cluster does not support the V2 APIs, only the normal MRS cluster can be interconnected. If the customer requires the MRS cluster with Kerberos authentication enabled, you are advised to use the DAYU platform for interconnection.

**NOTE**

- For MRS 1.X clusters, MRS 1.8.7 or later supports the V2 APIs.
- For MRS 2.X clusters, MRS 2.0.3 or later supports the V2 APIs.

11.10 Why Is the MRS Cluster with Kerberos Authentication Enabled Not Found When a Connection Is Set Up in DLF?

DLF cannot interconnect with MRS clusters with Kerberos authentication enabled. If you want to use MRS clusters with Kerberos authentication enabled, you are advised to use the DAYU platform.
11.11 How Do I Use PySpark on an ECS to Connect to an MRS Spark Cluster with Kerberos Authentication Enabled on the Intranet?

Change the value of `spark.yarn.security.credentials.hbase.enabled` in the `spark-defaults.conf` file of Spark to `true` and use `spark-submit --master yarn --keytab keytabfile --principal principal` to specify the Kerberos authentication file.

11.12 Why Is There No Mapping Field When HBase Synchronizes Data to CSS?

After data is synchronized from HBase of MRS to Cloud Search Service (CSS), the entire database does not have mapped fields. Only a single table has mapped fields.

11.13 Can a Spark Cluster Access Data in OBS?

The Spark cluster can access data stored in the OBS system. For details, see Configuring a Cluster with Storage and Compute Separated.

11.14 What Are the Differences Between Spark in MRS and Spark in DLI?

- The Spark of MRS is deployed on the VM allocated by the MRS you purchased. You can adjust and optimize the Spark service based on the site requirements. Various APIs can be called. For details, see Spark APIs.
- The Spark in Data Lake Insight (DLI) is a fully managed service. Customers are unaware of the Spark component and can only use this service. The APIs are encapsulated. For details, see JDBC API Reference.

11.15 How Can I Import Data from OBS to HDFS in Flume?

To import data from OBS to HDFS in Flume, you need to load the `hdfs-site.xml` and `core-site.xml` files to a Flume server and restart Flume.

11.16 Does the Component Interconnection Support the External KDC?

MRS supports only the built-in key distribution center (KDC).
11.17 When the Open Source Kylin 3.x Is Interconnected with the MRS 1.9.3 Version, the Jetty Version Compatibility Problem Exists

For security purposes, MRS upgrades some open source third-party components that have serious security vulnerabilities. As a result, the Jetty version compatibility problem occurs when the open-source Kylin interconnects with MRS 1.9.3.

Perform the following operations to complete the deployment and interconnection.

1. Install the MRS client on an ECS node. For details, see [Using an MRS Client on Nodes Outside a Cluster](#).
2. After the installation is complete, run the following command to import environment variables:
   ```bash
   source ./bigdata_envexport
   HIVE_CONF=/srv/client/Hive/config/export
   HCAT_HOME=/srv/client/Hive/HCatalog
   ```
3. Install Kylin on the node where the MRS client is installed and specify KYLIN_HOME. For details, see the [Kylin official website](#). For MRS 1.9.3, select Kylin for HBase 1.x for interconnection.
   ```bash
   export KYLIN_HOME=/srv/client/apache-kylin-3.0.2-bin-hbase1x
   ```
4. Remove Jetty-related JAR packages from the /srv/client/Hive/Beeline/lib/ directory in the Hive client directory to avoid version conflicts.
   Jetty JAR packages:
   ```
   javax-websocket-server-impl-9.4.26.v20200117.jar
   websocket-server-9.4.26.v20200117.jar
   jetty-all-9.4.26.v20200117-uber.jar
   jetty-runner-9.4.26.v20200117.jar
   apache-jsp-9.4.26.v20200117.jar
   ```
5. Start the Kylin service and check the Kylin logs. In normal cases, the logs do not contain compatibility exceptions such as `java.lang.NoSuchMethodException` or `java.lang.ClassNotFoundException`.
   ```bash
   SKYLIN_HOME/bin/kylin.sh start
   ```
6. Access the native Kylin page at [http://<hostname>:7070/kylin](http://<hostname>:7070/kylin) and run the sample Cube script `S(KYLIN_HOME)/bin/sample.sh` to check whether Kylin is running properly.
12.1 Can I Switch Between the two Login Modes of MRS?

No. The login mode cannot be changed after it is selected during cluster purchase.

12.2 How Do I Remotely Use an MRS Cluster?

For details about how to remotely log in to an MRS cluster, see Remote Operation Guide.

12.3 How Do I Set the IP Address and Port Number of the MRS Cluster for Local Joint Commissioning?

You can bind EIPs to all nodes that need to log in to the cluster. For details, see Assigning an EIP and Binding It to an ECS.

You can configure the port based on service requirements. For details, see List of Open Source Component Ports.

12.4 How Do I Obtain the ZooKeeper Address?

- For MRS 2.0.1 or later, perform the following steps:
  a. On the MRS cluster details page, click the Dashboard tab page, click Click to synchronize on the right side of IAM User Sync to synchronize IAM users.
Choose **Components > ZooKeeper > Instances** and obtain the **OM IP Address** of a ZooKeeper instance.
For MRS 2.0.1 or later, perform the following steps:

a. Log in to Accessing MRS Manager.

b. Choose Services > ZooKeeper > Instance and obtain the OM IP Address of a ZooKeeper instance.
13.1 What Are the Differences Between Sample Project Building and Application Development in MapReduce? Is Python Code Supported?

1. The sample project and application development in MapReduce are the same. You can select either of them.
2. MRS supports Python code.

13.2 What Compression Algorithms Does HBase Support?

HBase supports the Snappy, LZ4, and GZ compression algorithms.

13.3 How Do I Use HBase?

For details about how to use HBase, see Using HBase from ScratchHBase Application Development Overview.

For details about the websites of open source components such as HBase, see Web UIs of Open Source Components.

13.4 Can MRS Write Data to HBase Through the HBase External Table of Hive?

Not supported. Hive on HBase supports only data query but does not support data modification.
13.5 How Do I View HBase Logs?

1. Log in to the Master node in a cluster.
2. Run the `su - omm` command to switch to user `omm`.
3. Run the `cd /var/log/Bigdata/hbase/` command to go to the `/var/log/Bigdata/hbase/` directory and view HBase logs.

13.6 How Do I Access HBase Shell in a Cluster with Kerberos Authentication Enabled?

You need to download the MRS client from MRS Manager and install the MRS client on the access node. For details, see Using an MRS Client on Nodes Outside a Cluster.

13.7 Which Number Type of DynamoDB Is Better in Hive Tables?

Hive supports `smallint`, which is recommended.

13.8 Can the Hive Driver Be Interconnected with DBCP2?

The Hive driver cannot be interconnected with the DBCP2 database connection pool. The DBCP2 database connection pool invokes the `isValid` method to check whether a connection is available. However, Hive directly throws an exception when implementing this method.

13.9 How Does User A View the Hive Table Created by User B?

User A creates a role that has the permission to access the table created by user B. For details about how to create a role, see Configuring Hive Permissions. Then bind the newly created role to user A and wait for about 5 minutes. The table created by user B can be accessed. For details about how to modify the role information bound to a user, see Modifying User Information.

13.10 How Do I Use IDEA on a Local Computer to Access Hive?

You can access the MRS cluster to perform operations on Hive. For details, see Running the JDBC Client and Viewing Results.
13.11 How Do I Commission a SparkSQL Job in the Windows Environment?

For details about how to configure and commission SparkSQL jobs, see Preparing an Environment.

13.12 How Do I Reset Kafka Data?

To delete Kafka topic information is to reset Kafka data. For details, see the following commands:

- To delete a topic, run the `kafka-topics.sh --delete --zookeeper IP address:Port/kafka --topic Topic name` command.
- To query all topics, run the `kafka-topics.sh --zookeeper 192.168.0.4:2181/kafka --list` command.

After the deletion command is executed, if the topic data is empty, the topic will be deleted immediately. If the topic data exists, the topic is marked for deletion and will be deleted by Kafka later.

13.13 Can MRS Run Multiple Flume Tasks at the Same Time?

The Flume client can contain multiple independent data flows. That is, multiple sources, channels, and sinks can be configured in the `properties.properties` configuration file. These components can be linked to form multiple flows.

For example, to configure two data flows in a configuration, run the following commands:

```plaintext
server.sources = source1 source2
server.sinks = sink1 sink2
server.channels = channel1 channel2

#dataflow1
server.sources.source1.channels = channel1
server.sinks.sink1.channel = channel1

#dataflow2
server.sources.source2.channels = channel2
server.sinks.sink2.channel = channel2
```

13.14 How Do I Change FlumeClient Logs to Standard Output Logs?

1. Log in to the node where the Flume client resides.
2. Go to the Flume client installation directory. Assume that the Flume client installation directory is `/opt/FlumeClient`. Run the following command:

   ```bash
   cd /opt/FlumeClient/fusioninsight-flume-1.6.0/bin
   ```
3. Run the `.flume-manage.sh stop` command to stop Flume running.
4. Run the `vi /log4j.properties` command to open the `log4j.properties` file and change the value of `flume.root.logger` to `${flume.log.level},console`.
5. Run the `vim /flume-manager.sh` command to open the `flume-manager.sh` script in the `bin` directory in the Flume installation directory.
6. Comment out the following information in the `flume-manager.sh` script:
   ```
   >/dev/null 2>&1 &
   ```
7. Run the `.flume-manage.sh start` command to restart FlumeClient.
8. After the modification, check whether the docker configuration is correct.

### 13.15 How Does HDFS Balance Data?

1. Log in to the Master node of the cluster and run the following command to start the balancer:
   ```
   source /opt/client/bigdata_env
   sh /opt/client/HDFS/hadoop/sbin/start-balancer.sh -threshold 5
   ```
2. View the log.
   After you execute the balance task, the `hadoop-root-balancer-Host name.out` log file is generated in the client installation directory `/opt/client/HDFS/hadoop/logs`.
3. (Optional) If you do not want to perform data balancing, run the following command to stop the balancer:
   ```
   source /opt/client/bigdata_env
   sh /opt/client/HDFS/hadoop/sbin/stop-balancer.sh -threshold 5
   ```

### 13.16 How Do I Change the Number of HDFS Replicas?

1. Log in to MRS Manager.
2. Choose Services > HDFS > Service Configuration.
3. Set Type to All and search for `dfs.replication`.
4. Change the parameter value to a value ranging from 1 to 16, click Save Configuration, and restart the HDFS instance.

### 13.17 How Do I Use Python to Remotely Connect to an HDFS Port?

The default port of open source HDFS is 50070 for versions earlier than 3.0.0 and is 9870 for version 3.0.0 or later. For details about the port numbers of open source components, see **List of Open Source Component Ports**.
13.18 Where Are the JAR Files and Environment Variables of Hadoop Located?

The hadoopstreaming.jar file is stored in the /opt/share/hadoop-streaming-* directory.

NOTE

The variable * depends on the Hadoop version.

The environment variables of JDK are located in the /opt/client/JDK/component_env directory.

The environment variables of Hadoop are stored in the /opt/client/HDFS/component_env directory.

The Hadoop client path is /opt/client/HDFS/hadoop.

13.19 What Compression Algorithms Does Kudu Support?

Currently, compression algorithms supported by Kudu include Snappy, LZ4, and zlib. The default value is LZ4.

13.20 How Do I View Kudu Logs?

1. Log in to the Master node in a cluster.
2. Run the su - omm command to switch to user omm.
3. Run the cd /var/log/Bigdata/kudu/ command to go to the /var/log/Bigdata/kudu/ directory and view Kudu logs.

13.21 Does OpenTSDB Support Python APIs?

OpenTSDB provides HTTP-based RESTful APIs that are language-independent. Any language that supports HTTP requests can interconnect to OpenTSDB. Therefore, OpenTSDB supports Python APIs.

13.22 How Do I View Container Logs of Yarn?

For details about the Yarn tasks started in MRS and container log directory, see Introduction to Logs.
13.23 How Do I Specify a Log Path When Submitting a Task in an MRS Storm Cluster?

You can modify the `/opt/Bigdata/MRS_XXX/1_XX_Supervisor/etc/worker.xml` file on the streaming Core node of MRS as required, set the value of `filename` to the required path, and restart the corresponding instance on MRS Manager.

You are advised not to modify the default log configuration of MRS. Otherwise, the log system may be abnormal.

13.24 Where Can I Download the Dependency Package (com.huawei.gaussc10) in the Hive Sample Project?

Currently, MRS does not have the `com.huawei.gaussc10` dependency package, which is a GaussDB dependency package and does not need to be configured. You are advised to exclude this package when building a Maven project.

13.25 Does HDFS Support OBS?

HDFS supports OBS. For details about the configuration and usage, see Accessing OBS Using obs.

13.26 Why Does the Kafka Access Protocol Configuration Not Take Effect?

For details, see "Brief Introduction to API for Kafka" in Safety Instructions on Using Kafka. The configuration takes effect only after you configure the protocol based on user requirements.

13.27 How Do I Set and Modify the Data Retention Period of an HBase Table?

- Setting the retention period when a table is created:
  Create the `t_task_log` table, set the column family to `f`, and set the TTL to 86400 seconds.
  ```
  create 't_task_log',{NAME => 'f', TTL=>'86400'}
  ```

- Setting the retention period based on the existing table:
  ```
  disable "t_task_log" #Disable the table (services must be stopped).
  alter "t_task_log",NAME=>'data',TTL=>'86400' # Set the TTL value, which applies to the column family data.
  enable "t_task_log" #Restore the table.
  ```
13.28 How Do I Specify an Application Name in the flink run Command?

You can specify an application name (Application-Name) by adding -ynm in the flink run command. For details about how to submit a Flink job, see Running a Flink Job.

13.29 How Do I Connect to HBase of MRS Through HappyBase?

ThriftServer1 and ThriftServer2 cannot coexist. The HBase service of the MRS cluster uses ThriftServer2. However, HappyBase can connect to HBase only through the ThriftServer1 interface. Therefore, you are advised to use Python to directly connect to HBase, for details, see https://github.com/huaweicloud/huaweicloud-mrs-example/blob/mrs-1.8/src/hbase-examples/hbase-python-example/DemoClient.py.

13.30 How Do I Change the Resource Ratio of Yarn Queues?

For details about how to change the resource ratio of Yarn queues, see Configuring the Queue Capacity Policy of a Resource Pool.

13.31 How Do I Specify a Flink Task to Run in a Specified Yarn Queue?

You can use -qu, --queue <arg> to specify a Yarn queue when submitting a task. For details about the flink run parameter, see Flink Shell Commands.

13.32 How DO I View the Client Version of MRS Kafka?

Run the --bootstrap-server command to view the information about the client of the new version.

13.33 What Access Protocols Are Supported by Kafka?

The access protocols include PLAINTEST, SSL, SASL_PLAINTEXT, SASL_SSL.

13.34 How Do I Configure Other Data Sources on Presto?

To connect Presto to the self-built MySQL database, configure the connector file in the /opt/Bigdata/MRS_XXX/1_21_Coordinator/etc/catalog directory.
13.35 Does Hive Query Data Can Be Exported?

Hive query data can be exported by running the following statement:

```
insert overwrite local directory "/tmp/out/" row format delimited fields terminated by "\t" select * from table;
```

13.36 How Does MRS Connect to Impala Shell?

You can use the `impla-shell -i <host:port>` option. To automatically connect to a specific Impala database, use the `-d <database>` option. For example, if all your Kudu tables are in the `impala_kudu` database, `-d impala_kudu` can use this database. To exit Impala shell, use the `quit` command. For details, see Using Impala from Scratch.

13.37 How Does MRS Connect to Spark Shell?

1. Log in to the Master node in the cluster as user `root`.
2. Run the following command to configure environment variables:

   ```
   source /opt/client/bigdata_env
   ```
3. If Kerberos authentication is enabled for the current cluster, run the following command to authenticate the user. If Kerberos authentication is disabled, skip this step.
   ```
   kinit MRS cluster user
   ```
   Example:
   - If the development user is a machine-machine user, run `kinit -kt user.keytab sparkuser`.
   - If the development user is a human-machine user, run `kinit sparkuser`.
4. Run the client command of the Spark component:

   ```
   spark-shell
   ```

13.38 How Does MRS Connect to Spark Beeline?

1. Log in to the Master node in the cluster as user `root`.
2. Run the following command to configure environment variables:

   ```
   source /opt/client/bigdata_env
   ```
3. If Kerberos authentication is enabled for the current cluster, run the following command to authenticate the user. If Kerberos authentication is disabled, skip this step.
   ```
   kinit MRS cluster user
   ```
   Example:
   - If the development user is a machine-machine user, run `kinit -kt user.keytab sparkuser`.
   - If the development user is a human-machine user, run `kinit sparkuser`.
4. Run the client command of the Spark component:
   spark-beeline

5. Run the following command in spark-beeline. For example, create table test in the obs://mrs-word001/table/ directory.
   create table test(id int) location 'obs://mrs-word001/table/';

6. Run the following command to query all tables. If table test is displayed in the command output, OBS access is successful.
   show tables;

   ![Figure 13-1 Returned table name](image)

7. Press Ctrl+C to exit the Spark beeline.

### 13.39 Where Are the Execution Logs of Spark Jobs Stored?

- Logs of unfinished Spark jobs are stored in the /srv/BigData/hadoop/data1/nm/containerlogs/ directory on the Core node.
- Logs of tasks completed by Spark jobs are stored in the /tmp/logs/Username/logs directory of HDFS.
14.1 How Do I Configure the node_id Parameter When Using the API for Adjusting Cluster Nodes?

When you use the API for adjusting cluster nodes, the value of node_id is fixed to node_orderadd.
15 Cluster Management

15.1 How Do I View All Clusters?

You can view all MRS clusters on the Clusters page. If massive clusters are involved, you can turn pages to view clusters.

- **Active Clusters**: contain all clusters except the clusters in the Terminated state.
- **Cluster History**: contains the tasks in the Terminated and Failed state. Only clusters terminated within the last six months are displayed. If you want to view clusters terminated six months ago, contact technical support engineers.
- **Failed Tasks**: includes only tasks in Failed state. The failed tasks are as follows:
  - Clusters failed to be created
  - Clusters failed to be terminated
  - Clusters failed to be scaled out
  - Clusters failed to be scaled in

15.2 How Do I View Log Information?

You can view operation logs of clusters and jobs on the Operation Logs page. Currently, MRS has two types of operation logs:

- **Cluster operations**
  - Creating, terminating, scaling out, and scaling in a cluster
  - Creating and deleting a directory, and deleting a file
- **Job operations**: Creating, stopping, and deleting a job

*Figure 15-1* shows the operation logs.
15.3 How Do I View Configuration Information of a Cluster?

- After a cluster has been created, click the cluster name on the MRS management console to go to the **Basic Information** page. You can view the basic configuration information about the cluster, including the name, ID, billing mode, AZ, creation time, Hadoop component version, as well as the instance specifications and capacity of node. The data analysis and processing capability of the cluster depends on the instance specifications and capacity of nodes. More advanced instance specifications and larger capacity allow faster cluster running and better data processing, and accordingly requires higher cluster cost.

- On the **Basic Information** page, click **Manage** to access MRS Manager. On the MRS cluster management page, you can view and handle alarms, modify cluster configurations, and upgrade cluster patches.

15.4 How Do I Install Streaming Cluster Components in an MRS Analysis Cluster?

New components cannot be installed in a created cluster. To install streaming cluster components, you need to create a streaming cluster or hybrid cluster and select required components.

15.5 How do I Install Kafka and Flume in an MRS Cluster?

New components cannot be installed in a purchased cluster. Kafka and Flume are components of a streaming cluster. To install Kafka and Flume, you need to purchase a streaming or hybrid cluster or hybrid cluster, and select Kafka and Flume.

For details about how to use Kafka and Flume, see [Using Kafka](#) and [Using Flume](#).
15.6 How Do I Stop an MRS Cluster?

To stop an MRS cluster, click the name of each node on the Nodes tab page to go to the Elastic Cloud Server page and click Stop.

15.7 Do I Need to Shut Down a Master Node Before Scaling Up Its Specifications?

When the specifications of a Master node in the MRS cluster are scaled up, the Master node automatically shuts down. After the scale-up is complete, the system automatically starts.

15.8 Can the Capacity of a Data Disk Be Expanded on MRS?

MRS supports data disk capacity expansion. You are advised to expand the capacity during off-peak hours.

15.9 How Do I Add Components to an Existing Cluster?

Components cannot be added to or removed from an created MRS cluster. You are advised to create an MRS cluster that contains required components.

15.10 Can the Components Installed in an MRS Cluster Be Deleted?

The components in the MRS cluster cannot be deleted. If a component is not required, log in to MRS Manager and stop the component on the Services page.

15.11 Can the Node Type of an MRS Cluster Be Changed?

The node type of an MRS cluster cannot be changed. However, clusters of MRS 1.8.2 or later support Scaling Up Master Node Specifications.

15.12 Can MRS Cluster Nodes Be Changed on the MRS Management Console?

MRS cluster nodes cannot be changed on the MRS management console. You are not advised to change MRS cluster nodes on the ECS management console either. If you manually stop or delete an ECS, modify or reinstall the ECS OS, or modify ECS specifications for a cluster node on the ECS management console, the stability of cluster running may be affected.
If the preceding operations are performed on a node of the MRS cluster, MRS automatically identifies and deletes the node. You can log in to the MRS management console and restore the deleted node through scale-out. Do not perform operations on a node that is being added.

15.13 Can Only Core and Task Nodes with the Same Configurations as Existing Cluster Nodes Be Added During Cluster Scale-out?

During cluster scale-out, only nodes with the same configuration as existing cluster nodes can be added. When the corresponding nodes are sold out, you can add nodes with other configurations. For details about how to scale out a cluster, see Manual Scaling Out a Cluster.

15.14 Can the MRS Cluster in Use Be Expanded?

If the cluster in use is not faulty, you can directly scale out the cluster. For details, see Manual Scaling Out a Cluster.

15.15 How Do I Cancel Cluster Risk Notification?

1. Log in to the MRS management console.
2. Click the name of the cluster to enter its details page.
3. Choose Alarms > Notification Rules.
4. Locate the row that contains the rule to be modified, click Edit in the Operation column, and modify the subscription rule.
5. Click OK.

15.16 How Do I Configure Parameters for Bootstrap Actions?

For details about how to add a bootstrap action, see Adding a Bootstrap Action. For details about the script example, see Sample Scripts.
16 Kerberos Usage

16.1 Can I Change the Kerberos Authentication Status of a Purchased MRS Cluster?

The Kerberos service cannot be enabled or disabled after an MRS cluster is created. You need to enable or disable the Kerberos service when creating the MRS cluster. You are advised to create a cluster again.

16.2 How Do I Access MRS Manager That Supports Kerberos Authentication?

For details about how to access MRS Manager that supports Kerberos authentication, see Accessing MRS Manager Supporting Kerberos Authentication.

16.3 What Are the Ports of the Kerberos Authentication Service?

The Kerberos authentication service uses ports 21730 (TCP), 21731 (TCP/UDP), and 21732 (TCP/UDP).

16.4 How Do I Deploy the Kerberos Service in a Running Cluster?

The MRS cluster does not support customized Kerberos installation and deployment, and the Kerberos authentication cannot be set up between components. Therefore, you are advised to purchase another cluster with Kerberos enabled and migrate data.
16.5 How Do I Access Hive in a Cluster with Kerberos Authentication Enabled?

1. Log in to the Master node in the cluster as user root.
2. Run the following command to configure environment variables:
   ```bash
   source /opt/client/bigdata_env
   ```
3. If Kerberos authentication is enabled for the current cluster, run the following command to authenticate the user. The current user must have the permission to create Hive tables. For details about how to configure Hive permissions, see Configuring Hive Permissions, and bind roles to the user. If Kerberos authentication is disabled, skip this step.
   ```bash
   kinit MRS cluster user
   ```
   Example: `kinit hiveuser`
4. Run the client command of the Hive component.
   ```bash
   beeline
   ```
5. Run the Hive command in beeline. For example:
   ```sql
   create table test_obs(a int, b string) row format delimited fields terminated by "," stored as textfile location "obs://test_obs";
   ```
6. Press Ctrl+C to exit the Hive beeline.

16.6 How Do I Access Presto in a Cluster with Kerberos Authentication Enabled?

1. Log in to the Master node in the cluster as user root.
2. Run the following command to configure environment variables:
   ```bash
   source /opt/client/bigdata_env
   ```
3. Access Presto in a security cluster with Kerberos authentication enabled.
   a. Log in to MRS Manager and create a role with the Hive Admin Privilege permission, for example, `prestorerole`. For details about how to create a role, see Creating a Role.
   b. Create a user that belongs to the Presto and Hive groups, and bind the user to the role created in 3.a, for example, `presto001`. For details about how to create a user, see Creating a User.
   c. Authenticate the current user.
      ```bash
      kinit presto001
      ```
   d. On MRS Manager, choose System > Manage User. In the row of the new user, choose More > Download authentication credential.
e. Decompress the downloaded user credential file, and save the obtained `krb5.conf` and `user.keytab` files to the client directory, for example, `/opt/client/Presto/`.

f. Run the following command to obtain a user principal:

   ```bash
eklist -kt /opt/client/Presto/user.keytab
   ```

   For clusters with Kerberos authentication enabled, run the following command to connect to the Presto Server of the cluster:

   ```bash
   presto_cli.sh --krb5-config-path {krb5.conf file path} --krb5-principal {User's principal} --krb5-keytab-path {user.keytab file path} --user {presto username}
   ```

   - `krb5.conf file path`: Replace it with the file path set in 3.e, for example, `/opt/client/Presto/krb5.conf`.
   - `user.keytab file path`: Replace it with the file path set in 3.e, for example, `/opt/client/Presto/user.keytab`.
   - `User's principal`: Replace it with the result returned in 3.f.
   - `presto username`: Replace it with the name of the user created in 3.b, for example, `presto001`.

Example:

   ```bash
   presto_cli.sh --krb5-config-path /opt/client/Presto/krb5.conf --krb5-principal presto001@xxx_xxx_xxx_xxx.COM --krb5-keytab-path /opt/client/Presto/user.keytab --user presto001
   ```

h. On the Presto client, run the following statement to create a schema:

   ```sql
   CREATE SCHEMA hive.demo01 WITH (location = 'obs://presto-demo001/');
   ```

i. Create a table in the schema. The table data is stored in the OBS bucket, as shown in the following example:

   ```sql
   CREATE TABLE hive.demo01.demo_table WITH (format = 'ORC') AS SELECT * FROM tpch.sf1.customer;
   ```

j. Run `exit` to exit the client.
17 Metadata Management

17.1 Where Can I View Hive Metadata?

Hive metadata is stored in GaussDB of an MRS cluster. You can log in to the Master node of the cluster, switch to user omm, and run the `gsql -p {PORT} -U {USER} -W {PASSWD} -d hivemeta` command to view the metadata.